

A Hitchhikers Guide to the Black Arts (of Earth system modelling)

I: cGENIE Earth system model basics

Stuff to keep in mind:

- cGENIE is a model. Models ARE NOT the 'real World'. (Don't get confused!)
- The very low (for a 3-D ocean circulation model) resolution of cGENIE limits its applicability to very short time-scales, and because the simplified atmospheric component cannot calculate winds, there is no atmospheric dynamics or inter-annual variability.
- cGENIE is best thought of as a 'discovery and exploring' tool for learning how the Earth system works rather than a detailed 'simulation' tool.
- Have fun (or at least try) ☺

Relevant reading:

(1) Primary model description + evaluation/discussion papers:

Edwards and Marsh [2005] (Climate Dynamics 24, 415-433)

→ description and calibration of the climate model component of (c)GENIE

Hargreaves et al. [2004] (Climate Dynamics 23, 745-760)

→ description of data assimilation methodology and calibrated climatology of the climate model

Ridgwell et al. [2007a] (Biogeosciences 4, 87-104)

→ description of basic ocean carbon cycle and its calibration against observations

Ridgwell et al. [2007b] (Biogeosciences 4, 481-492)

→ description of the calibration and application of the pH-responsive pelagic calcification

Ridgwell and Hargreaves [2007] (Global Biogeochemical Cycles 21, doi:10.1029/2006GB002764)

→ description and calibration of the sediment model component and response to fossil CO₂ release

Lenton et al. [2006] (Climate Dynamics 26, doi:10.1007/s00382-006-0109-9)

→ description, calibration, and application of the 'ENTS' land surface + vegetation scheme

(2) Some potted model applications (mostly paleo):

Chikamoto, M. O., K. Matsumoto, and A. Ridgwell [2008] (JGR 113, doi:10.1029/2007JG000666)

→ deep-sea CaCO₃ sediment and atmospheric CO₂ response to AMOC shutdown

Panchuk, K., A. Ridgwell, and L. R. Kump [2008] (Geology 36, 315-318)

→ configuration of (c)GENIE for Palaeogene marine carbon cycling; assessing PETM CO₂ release

Ridgwell [2007] (Paleoceanography 22, doi:10.1029/2006PA001372)

→ description of sediment core modelling; application to the interpretation of PETM CCD changes

Singaraye, J. S., D. A. Richards A. Ridgwell, P. J. Valdes, W. E. N. Austin, and J. W. Beck, [2008] (GRL 35, doi:10.1029/2008GL034074)

→ analysis of ocean circulation and atmospheric radiocarbon variability during the Younger Dryas

Copies of these references can be obtained from the 'usual places' (i.e., 'journals!'), or from:

www.seao2.info/pubs.html or www.genie.ac.uk/publications/papers.html

Other cGENIE resources can be found at:

<http://mycgenie.seao2.info>

Andy (andy@seao2.org)

0. Before anything else ...

- 0.0 For some of you, the mechanics of running the model will be about as much fun as sticking your tongue in an electrical outlet (a popular hobby in England). (*However, if you are an experienced linux/unix/tounge-in-electrical-socket user, you can skip on to **Section 0.1.***)

cGENIE has traditionally been configured and accessed at the 'command line' of the linux (or Mac equivalent) operating system. The command line is a place where you type text and when you press Return, something (hopefully, good!) happens. Typically the stuff you type started with a 'Command' word, and often followed by one or more options and parameters. The Command word and any options / parameters MUST be separated by SPACES.

The command line is indicated with something like: \$. The \$ is called the 'prompt' and is 'prompting' you to type some input (Commands, swear words, etc.). See – the computer is just sat there waiting for you to command it to go do something (stupid?!). Often, you will also be informed (reminded) of the username and computer name and current directory at the prompt, e.g.:

```
[username@iwan ~]$
```

Which in this example is user 'username' (yours will be different) on computer 'iwan' and the current directory is the 'home' directory (~).

If you are not or not very familiar with the linux/unix command line – Appendix 1 contains a quick summary of some of the more important/useful Commands you can use at the command line.

NOTE: In the following workshop instructions, example lines to be typed at the command line **are highlighted in yellow** (or grey when photocopied ...). Be VERY CAREFUL that spaces are not missed out. Also be careful not to confuse the number one (1) for the letter el (l). Misspelling/typing will probably be the primary reason for any wailing and gnashing of teeth ...

NOTE-the-second: **Text highlighted in grey** represents instructions specific to using Windoz PCs, and hence may or will differ on different computers and/or under operating systems. Other software choices (e.g., for editing files or viewing results) and platforms (e.g. Macs) are possible.

- 0.1 IF YOU 'RACE AHEAD' AND SKIP OVER BITS THAT YOU CAN'T BE BOTHERED TO READ OR DO ... IT WILL ALL GO HORRIBLY PEAR-SHAPED BEFORE YOU CAN SAY "WHAT MUPPET WROTE THESE STUPID INSTRUCTIONS?".
- 0.2 Some warnings and reminders are repeated over and over and over and ... over again. This is because you will forget immediately each time! ;)
- 0.3 **OK – now we are ready to start ...**

1. Starting (dozing?) off ...

- 1.1 Log in to the account that has been created for you on the almond computing cluster. To do this – first start the SSH Secure Shell program (the Secure Shell Client). Click on the Quick Connect icon in the main Secure Shell Client window. A 'Connect to Remote Host' dialogue box will appear. In the first box, 'Host Name', enter:
`iwana.ggy.bris.ac.uk`
and your computing cluster user-name on the line below this ('User Name').
When (and only when) you click on the Connect button will you be asked for your password in a new dialogue window that will appear.
You should now have a dull, blank-look window with the command prompt (\$) (see previous page).
For displaying directory contents and transferring model results files we are going to be using the SSH file transfer program. One of the icons towards the middle of the icon bar in the SSH terminal window you have opened is called 'New File Transfer Window' (it should be the 2nd icon to the right of the search (binoculars) icon). Click on this to open a file transfer window.
You should now have TWO windows open – a 'shell' window (lines of text on an otherwise blank screen) and a file manager (transfer) window. Ensure that you have both these before moving on. It is recommended that you maximize both these windows to full screen. (But no-one will die horribly for not doing so. Probably.)
- 1.2 Carry out a brief check that cGENIE is installed OK:
(1) Change directory (see: Appendix I) to `~/cgenie/genie-main` and type:
`$ make testbiogem`
This compiles a carbon cycle enabled configuration of cGENIE and runs a short test, comparing the results against those of a pre-run experiment.
There may be some 'Warnings' reported (== someone's sloppy programming) but these are not detrimental to the ultimate science results (we hope!). 'Success' of this test is indicated by:
`**TEST OK**`
(2) At this point, the science modules are currently compiled in a grid and/or number of tracers configuration that is unlikely to be what you want for running experiments. Clean up (remove) all the compiled cGENIE modules, ready for re-compiling afresh from the source code by typing:
`$ make cleanall`
- 1.3 Later on you will be editing some configuration files. So now might be a good time to check that you can use the editor! (You will also be using the same editor to view some of the model output.)
You have two alternative options for editing and viewing text files, depending on whether you are a UNIX nerd with no life, or prefer anything to do with computers to be wrapped in cotton wool and covered with dollops of treacle.
EITHER: Use the linux `vi` application (or similar) if you are familiar with it. I think that this pretty much sucks as a text editor and life is far too short and brutal ... so I will not *make* you ;)
OR ... Use a suitable linux-friendly text editor (NOT Micro\$oft Notepad) in conjunction with the Secure File Transfer Client. For example: 'SciTE'.
To set SciTE to automatically open the model configuration files: in the Secure File Transfer Client – go to Edit, then Settings... and from the SECOND File Transfer section of the list in the left-hand panel, click on the button for If a file association is missing. Then find the program executable (`SciTE.exe`) and click OK. Under File in the main menu: Save Settings.
It should now be possible to double-click on a file in the Secure File Transfer Client and it will open like magic (almost)! Try opening `runcgenie.sh` (directory: `cgenie/genie-main`) in this way. If you edit and save the file, you will be asked whether you want to transfer it back to the remote machine and also whether you want to over-write the original file. Simply click Yes to both.

2. Running the model ('interactively', in a shell window)

2.1 The strategy for running the model is as follows:

At the command-line (\$) in the `genie-main` directory (not your home directory), you will be entering in a command (`runcgenie.sh`) together with a list of parameters that will be passed to the model, and as if by magic the model will run (or sometimes not):

```
$ ./runcgenie.sh #1 #2 #3 #4 (#5)
```

You must list at least 4 parameters after `./runcgenie.sh`, separated by S P A C E S and on a single continuous line (even if it 'wraps' around across 2 lines of the screen). For example:

```
$ ./runcgenie.sh cgenie_eb_go_gs_ac_bg.worbe2.ANTH LABS expl_modern 10
parameter number =  ↑ #1                                ↑ #2 ↑ #3                                ↑ #4
```

(Take care not to confuse an el ('l') with a one ('1') when typing this in ... (it is a 'one' here).)

These parameters are:

#1 ... is the name of the required base (or 'basic') configuration ('*base-config*') of the model.

Initially, we will mostly be using the base config: **cgenie_eb_go_gs_ac_bg.worbe2.ANTH**, which specifies the cGENIE climate model components: GOLDSTEIN ocean (go) + (GOLDSTEIN) sea-ice (gs) + EMBM atmosphere (eb) together with ocean (bg) and atmosphere (ac) carbon cycle modules (and a FULL-ish selection of ocean geochemical tracers). The ocean circulation component is configured with 8 vertical levels and is forced with annual average insolation. This configuration is basically as described in *Ridgwell et al. [2007]*.

#2 ... is the name of the subdirectory (if any) containing the user configuration ('*user-config*') file (i.e., the file containing the specification of a particular experiment). In the particular file structure adopted here, it will be: **LABS**. (If a *user-config* to be used does not live in a subdirectory of `cgenie/genie-userconfigs`, then a '/' is passed as this parameter (a blank or space will be hideously mis-interpreted ...).)

#3 ... is the name of the experiment itself. There must be a file in the directory specified by parameter **#2** (`cgenie/genie_userconfigs/LABS`) with exactly the same name as you enter here for parameter **#3**. i.e. the actual experiment is defined by the *user-config* file. Here, the experiment is called: **expl1_modern**.

#4 ... is the run length of the experiment in years – this must be entered as an integer (even though cGENIE will actually be treating it as a real number). In this example, we are running for only **10** years.

There is also one optional (5th) parameter (not used in this particular example and so left out (blank)):

#5 ... is the full path (and name) of any model experiment that you wish to continue on from the end of (called a '*restart*' file). If the 5th (optional) parameter is not passed then cGENIE will run from 'cold'. If the 5th parameter is present then cGENIE will attempt to run from a previously generated (*restart*) state. *Restarts* will be discussed later ...

2.2 What should happen is: First, you will end up twiddling your thumbs a while, as all the components of cGENIE are compiled from the raw source code (FORTRAN). When it has finished doing this, the model will initialize and carry out some brief self-checking. Only then will it start actually 'running' and doing something, starting with a header describing the columns of numbers that follow:

```
pCO2 (uatm)  – mean atmospheric CO2 concentration (in units of μatm)
d13CO2       – mean δ13C value of atmospheric CO2 (‰) (NOTE: only valid if 13C tracer is selected)
AMO (Sv)     – Atlantic meridional overturning circulation (Sv) (assuming modern continents)
ice (%)      – global sea-ice fraction (%)
<SST>        – global sea surface temperature 'SST' (°C)
<SSS>        – global sea surface salinity 'SSS' (‰)
<DIC>        – global mean ocean dissolved inorganic carbon (DIC) concentration (μmol kg-1)
<ALK>        – global mean ocean alkalinity (ALK) concentration (μeq kg-1)
```

(The choice of what information to display on screen as the model is running is rather arbitrary, but the chosen metrics do tend to summarize some of the main properties of the climate system and carbon cycle – for my own personal convenience rather than reflecting any fundamental scientific truth ...)

```
*****
*** Initialisation complete: simulation starting ...
*****
```

	model	year	* pCO2 (uatm)	d13CO2	* AMO (Sv)	ice (%)	<SST>	<SSS>	* <DIC> (uM)	<ALK> (uM)
\$N\$		0.00	278.000	-6.500	0.000	0.000	0.000	34.900	2244.000	2363.000
>>>	SAVING BIOGEM TIME-SLICE AVERAGE CENTERED @ year :						0.500			
>>>	SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :						0.500			
\$N\$		1.00	290.868	-7.075	13.613	0.744	2.508	34.901	2241.557	2363.109
>>>	SAVING BIOGEM TIME-SLICE AVERAGE CENTERED @ year :						1.500			
>>>	SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :						1.500			
\$N\$		2.00	299.214	-7.448	12.829	3.515	4.470	34.901	2240.416	2363.136
>>>	SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :						2.500			
\$N\$		3.00	305.330	-7.698	11.698	5.138	6.002	34.901	2239.653	2363.162
>>>	SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :						3.500			
\$N\$		4.00	309.895	-7.867	10.447	5.986	7.211	34.901	2239.102	2363.191
>>>	SAVING BIOGEM TIME-SLICE AVERAGE CENTERED @ year :						4.500			
>>>	SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :						4.500			
\$N\$		5.00	313.435	-7.982	9.384	6.215	8.155	34.902	2238.679	2363.221
>>>	SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :						5.500			
\$N\$		6.00	316.225	-8.060	8.505	6.641	8.973	34.902	2238.350	2363.246
>>>	SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :						6.500			
\$N\$		7.00	318.430	-8.111	7.926	6.644	9.635	34.903	2238.090	2363.268
>>>	SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :						7.500			
\$N\$		8.00	320.191	-8.141	7.921	6.750	10.223	34.903	2237.885	2363.286
>>>	SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :						8.500			
\$N\$		9.00	321.567	-8.156	7.956	6.752	10.730	34.904	2237.727	2363.302
>>>	SAVING BIOGEM TIME-SLICE AVERAGE CENTERED @ year :						9.500			
>>>	SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :						9.500			
\$N\$		10.00	322.625	-8.158	8.028	6.709	11.175	34.904	2237.609	2363.315
>>>	SAVING BIOGEM RESTART <<<									
>>>	SAVING ATCHEM RESTART <<<									

```
*****
*** Simulation complete: shutdown starting ...
*****
```

This information is reported at the same intervals as time-series data (see later & refer to the *User Manual*) is saved and is indicated by:

```
>>> SAVING BIOGEM TIME-SERIES AVERAGE CENTERED @ year :
```

Interleaved with these lines are lines reporting the saving of time-slice data (the 2- and 3-D model states – more of which later and in the *User Manual*). These appear as:

```
>>> SAVING BIOGEM TIME-SLICE AVERAGE CENTERED @ year:
```

When the model ends (as per the value you specified at the command line), a bunch more lines of diagnostics things are reported. If you scroll up the screen of the terminal window you can get back to where the time-dependent information appears while the model was running.

You can also stop the model at any point (all data up to that time will have been saved) by hitting: **<Ctrl-C>** (CONTROL key + 'C' key).

3. Model output

- 3.1 The first thing to note about output (i.e., saved results files) from cGENIE is that every science module saves its own results in its own sub-directory (and sometimes in very different and difficult-to-fathom ways ...). All the sub-directories of results, plus copies of input parameters and the model executable, are gathered together in a directory that is assigned the same name as the experiment (= *user-config* file name). The overall experiment result directory is located within:

```
~/cgenie_output
```

and will be assigned a directory name something like:

```
expl_modern
```

Within this are each module's results sub-directories.

We will consider primarily only results saved by the ocean biogeochemical module 'BIOGEM' (subdirectory: *biogem*). The results files will thus be found in:

```
~/cgenie_output/expl_modern/biogem
```

BIOGEM has a flexible and powerful facility of saving results by means of spatially explicit '*time-slices*', and as a semi-continuous '*time-series*' of a single global (or otherwise representative mean) variable. In contrast, ATCHEM does not save its own results (BIOGEM can save information about atmospheric composition and air-sea gas exchange) while SEDGEM essentially saves results only at the very end of a model experiment (BIOGEM can also save the spatial distribution of sediment composition as time-slices as well as mean composition as a time-series). Furthermore, in order to attain a common format for both ocean physical properties and biogeochemistry, BIOGEM can save a range of ocean results in addition to temperature and salinity, such as: velocities, sea-ice extent, mixed layer depth, convective frequency, etc

Time-slices

One of the most informative data sets that can be saved is that of the spatial distribution of properties (such as tracers or physical ocean attributes). However, saving full spatial distributions (a 36×36×8 array) for any or all of the tracers each and every time-step is clearly not practical; not only in terms of data storage but also because of the detrimental effect that repeated file access has on model run-time.

Instead, BIOGEM will save the full spatial distribution of tracer properties only at one or more predefined time points (in units of years). These are termed *time-slices*. At the specified time points, a set of spatially-explicit data fields are saved for all the key tracer, flux, and physical characteristics of the system. However, rather than taking an instantaneous snapshot, the *time-slice* is constructed as an average over a specified integration interval (the default is set to 1.0 years, i.e. an annual average).

Time-series

The second data format for model output is much more closely spaced in time. Model characteristics must then be reducible to a single meaningful variable for this to be practical (i.e., saving the time-varying nature of 3-D ocean tracer distributions is not). Suitable reduced indicators would be the total inventories in the ocean and/or atmosphere of various tracers (or equivalently, the mean global concentrations / partial pressures, respectively). Like the *time-slices*, the data values saved in the *time-series* files represent averages over a specified integration interval (the default is set to 1.0 years (annual average)).

- 3.2 The results directory will contain files with names of the form:

- (i) *biogem* (is the re-start file created from the run you have just complete, and can be ignored).
- (ii) *biogem_series_*.res* – these are the time-series files (in ASCII / plain text format).
- (iii) *biogem_year_*_diag_GLOBAL.res* – these contain (global diagnostics) summary information and are saved at the same frequency as the time-slices (also as ASCII / plain text).
- (iv) *fields_biogem_2d.nc* – 2-D fields of ocean and atmosphere properties, as NetCDF.
- (v) *fields_biogem_3d.nc* – 3-D fields of ocean properties, as NetCDF.

4. Viewing time-series output

- 4.1 A descriptive summary of all the time-series (biogem_series_*.res) data files is given in the cGENIE *User Manual* if you are really that bored. The files of most immediate use/relevance are:

biogem_series_atm_humidity.res	- mean atmospheric (surface) humidity
biogem_series_atm_temp.res	- mean atmospheric (surface) air temperature
biogem_series_misc_opsi.res	- mean ocean overturning stream-function; Atlantic/global min, max
biogem_series_misc_seaice.res	- mean ocean sea-ice cover and thickness
biogem_series_ocn_sal.res	- mean ocean surface and whole ocean salinity
biogem_series_ocn_temp.res	- mean ocean surface and whole ocean temperature

- 4.2 The contents of the files can be viewed by changing directory into the results directory, i.e.,:

```
$ cd ~/cgenie_output/expl_modern/biogem
```

and opening a file in the `vi` editor. But then I promised that I would not suggest anything quite so unpleasant ...

Instead – change to the experiment results directory and then to the BIOGEM sub-directory in the Secure File Transfer Client, and try double-clicking on one of the .res files (listed above). For `biogem_series_ocn_temp.res`, you should see 4 columns – time, mean (whole) ocean temperature (°C), mean (sea) surface temperature ‘SST’ (°C), and mean benthic temperature (°C). Other results files may differ in the numbers of columns but all should be identifiable from the header information.

Note: The Secure File Transfer Client **does not** automatically refresh the directory listing. If you cannot see the results sub-directory with the experiment name you have just run 99 times out of 100, it is because the display of the Secure File Transfer Client needs to be refreshed – there is an icon at the top of the program window or hit the ‘F5’ key.

- 4.3 For your information and edification: Excel, or MUTLAB if you prefer, can be used to graph the time-series results. Either way you will have to deal with the header line(s) that are present at the top of the file (and preceding the rows of data).

In Excel: Chose File then Open. You will want to select Files of Type ‘All Files (*.*)’. In the Text Import Wizard window you can request that Excel skips the first few lines to start the import on the 2nd or 3rd line of the text file. Alternatively, set appropriate column widths manually in Excel to ensure that the columns of data are correctly imported.

MUTLAB will ignore lines starting with a %, which the time-series starts with. However, it may be that the header line wraps-around and there is in effect a 2nd header line but without a %. In this case, extra care (or a quick edit of the header in the ASCII file) will be required to load the data into MUTLAB.

5. Viewing time-slice output of 2-D and 3-D environmental property fields

- 5.0 For the time-slice NetCDF (*.nc) files you will be using a program called 'Panoply'. If you want your own (FREE!) copy of this utility, you can get it here (and is available for: Windoz, Mac, and linux operating systems): <http://www.giss.nasa.gov/tools/panoply/>.
Alternatively: there is a Windoz copy sat on the cGENIE website (www.seao2.info/mycgenie.html).
- 5.1 The first time that you try and view a NetCDF file you will need to 'associate' the file format (the .nc file format extension) with the Panoply program. To do this: (1) copy across to a local directory either the 2-D or 3-D NetCDF results file from the BIOGEM results sub-directory, (2) right-mouse-button click over the file, and chose Open With..., (3) Browse ... and find the Panoply program file. Select and tick: Always use the selected program to open this kind of file. Hopefully ... all you have to do now is to double-click on the .nc file icon in the File Transfer window and the time-slice will be displayed like ... something that is *"indistinguishable from magic in a sufficiently retarded civilization"* (mis-adapted from A. C. Clarke).
- 5.2 When you open the NetCDF file, you will be presented with a 'Datasets and Variables' window (on the left hand side of the application window). This contains a list of all the parameters available that you can display. You will find that the 'Long Name' description of the variable will be the most helpful to identify the one you want. Simply double-click on a variable to display. For the 3-D fields you will be asked first whether you want a 'Lon-Lat' or 'Lat-Vert' plot (for the 2-D fields, the plot display will immediately open).
- For the 'Lon-Lat' plots – there are multiple levels (depth layers) in the ocean of data that can be plotted, from the surface to the abyssal ocean.
- For the 'Lat-Vert' plots – there are multiple possible longitudes at which to plot slices. The default is the global mean meridional distribution.
- For both: there may be multiple time-slices (i.e., you can plot data saved from different years). You can interpolate the data or not (often you may find that it is clearer not to interpolate the data but to leave it as 'blocky' colors corresponding to the resolution of the model), change the scale and colors, overlay continental outline, change the projection, etc etc. Gray cells represent 'dry' grid points, i.e., continental or oceanic crust.
- NOTE: The default settings in Panoply can mislead:
- (1) the 1st time-slice (often year mid-point 0.5) rather than the experiment end,
 - (2) auto-scaling of the color scale,
 - (3) global zonal averaging of lat-depth plots.
- Be careful when opening a new plot that you are looking at what you **think** you are looking at ...
- 5.3 Explore different data fields and play with different ways of displaying them. Aim for a set of display properties that show the information you are interested in / want to present in the clearest possible manner. Try different years (time-slice number), depth level (for a lat-long plot), or longitude (for a vertical section). Note that gray cells represent 'dry' grid points, i.e., continental or oceanic crust.
- 5.4 To save plots in Panoply:
- File
- Save Image As ...
- Then select the location, filename, and graphics format.
- 5.5 There is also a set of MUTLAB script for plotting NetCDF output in the directory:
- ~/cgenie/genie-matlab

6. Submitting experiment 'jobs' to a cluster

- 6.0 This bit is no particular fun at all, but it is a very handy 'trick' for running the model in the background, and maximizes drinking time in the pub vs. sat bored watching a computer screen ☺
- 6.1 Running jobs interactively is all very well, but there are three important limitations:
- (1) The connection between your terminal and the server computer running the model must remain unbroken. Anything more than a fleeting loss of internet connectivity may result in the experiment terminating.
 - (2) You can only run one experiment at a time ... unless you want to have thousands of separate X-Windows open ...? I thought not ...
 - (3) Any cluster or computer you are likely to be accessing using a shell will not have many computing cores, either because it is a single machine with only one or two processors, or if a cluster, by using an X-Window you are running on the 'head node', which will have similar computing core limitations to running on a single machine. The more experiments you run simultaneously, the slower they will all run ...
- 6.2 The alternative is to submit your experiment as a 'job' to a queuing system which then manages what compute resources are used to run the model. Once you have submitted the experiment, that is it – you can go straight to the pub :)

Unless you want to read the *cGENIE User Manual*, for now, take the following suggested way of submitting jobs on trust: to run an experiment (`exp2_modern`) for 10 years by submitting the experiment as a job to the cluster queue, type:

```
$ qsub -q cat.q -j y -o cgenie_log -S /bin/bash  
runcgenie.sh cgenie_eb_go_gs_ac_bg.worbe2.ANTH LABS exp2_modern 10
```

(Again: SINGLE, CONTINUOUS LINE.)

Note that now you should omit the `./` bit before `runcgenie.sh`.

(If you are interested (I know that you are not): the options following `qsub` and before `runcgenie.sh` do things like re-directing screen output and error messaging to a file and specify which linux 'shell' to assume. It is even possible to receive an email when the job is done ☺)

The status of the cluster queue and how your experiment job is getting on (e.g., "Is it finished yet?") can be checked by typing:

```
$ qstat -f
```

(`qstat -f -u "*" will show all jobs on the cluster.`)

After submitting an experiment, you receive a job number. This number appears in the first column in the queue status information when you issue a `qstat -f` command. You should see your job appear on one of 2 compute nodes of the 'cat' queue, numbered 0-2 through 0-9 (although there are 10 compute nodes in total on the full queue, numbered from 0-0 to 0-9), although it might briefly reside as a 'PENDING JOB'. For each node, there are 8 processing cores, meaning that up to 8 instances of cGENIE can run simultaneously on each node (i.e., a total of 80 across the entire cluster). For an 8-level ocean based configuration of cGENIE, being run for 10 years, the job should remain there in the queue for a few good tens of seconds before 'disappearing' (your clue that it has finished, or died ...). If you periodically re-issue a `qstat -f` command you can follow your job's progress.

A rough rule of thumb is that 8-level ocean cGENIE will simulate about 1000 years per CPU hours. The 16-level version (which you will use later), about 300-400 years per CPU hour.

NOTE: If your experiment appears on the queue but vanishes after a few seconds, it has most likely died ☹

7. 'Restarts'

7.0 Not much fun here either ... but again – an important and time-saving (== increased drinking time!) modelling technique to learn to use.

7.1 By default, model experiments start from 'cold', i.e., the ocean is at rest and uniform in temperature and salinity while the atmosphere is uniform in temperature and humidity. All biogeochemical tracers in the ocean have uniform concentrations and/or are zero and there are no biogenic materials in deep-sea sediments. From this state it will take several thousand years (kyr) for the climate system to reach steady-state, closer to 10 kyr for ocean biogeochemical cycles and atmosphere CO₂ to reach steady-state, and approaching 100 kyr for sediment composition to re-balance weathering. Reaching this the equilibrium state is called the 'spin-up' phase of the model. There is evidently little point in repeating the spin-up for each and every model experiment that might similar except in a single detail (e.g., testing a variety of different CO₂ emissions scenarios all starting from current year 2012 conditions). A facility is thus provided for requesting that a 'restart' is used – starting a new experiment from the end of a previous one, usually a 'spin-up' that has been run explicitly for the purpose of generating a starting point (*restart*) for subsequent experiments to continue on from.

7.2 A *restart* can be requested by setting the 5th (optional) parameter when entering in the `runcgenie.sh` command. A spin-up of the preindustrial Earth system, based on an 8-level resolution of the ocean model and near identical to that described in *Ridgwell et al. [2007a]* is provided: `exp0_modern_SPINUP` (which can be found in the `~/cgenie_output` directory). The experiment `exp3_modern` can be told to use a *restart* state by adding an appropriate 5th parameter at the command line, e.g.:

```
$ ./runcgenie.sh cgenie eb go gs ac_bg.worbe2.ANTH LABS  
exp3_modern 10 exp0_modern_SPINUP
```

The run-time output should now look noticeably different. There should be no (very little) drift in any of the various variable values outputted to the screen – this is because you have (re-)started from the end of a run that had already reached an equilibrium, steady-state.

Appendix 1 – linux 101

Throughout the cGENIE tutorials and documentation, your home directory will be represented by '~'. Thus, for example, the output directory path will be written:

```
~/cgenie_output
```

Viewing directories; moving around the file system

At the command line prompt (\$) in linux, you can view the current directory contents:

```
$ ls
```

or for a more complete output:

```
$ ls -la
```

To go down a directory (cgenie_output) is:

```
$ cd cgenie_output
```

and to go back up one is:

```
$ cd ..
```

You can always return to your home directory (~) by typing:

```
$ cd
```

or:

```
$ cd ~
```

Copying and moving files

To copy a file myconfig to myconfig_new:

```
$ cp myconfig myconfig_new
```

To move myconfig to the cGENIE user-config directory:

```
$ mv myconfig ~/cgenie/genie-userconfigs/LABS/myconfig
```

To rename myconfig to useless_config:

```
$ mv myconfig useless_config
```

Creating directories

To create a directory mydirectory:

```
$ mkdir mydirectory
```

Repeating command lines

You do not have to re-enter lines of commands and options in their entirety each time – by pressing the UP cursor key you get the last command you issued. If you keep pressing the UP cursor key you can recover progressively older commands you have previously entered. When you have recovered a helpful line you can simply just edit it, navigating along with the LEFT and RIGHT cursor keys (press RETURN when you are done).